MULTIMEDIA UNIVERSITY

STUDENT ID NO | | | | | | | | | | |

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2022/2023 (TERM 2215)

DSP5614 – PROBABILITY AND STATISTICS

(for diploma students only)

19 DECEMBER 2022

9.00 am – 11.00 am

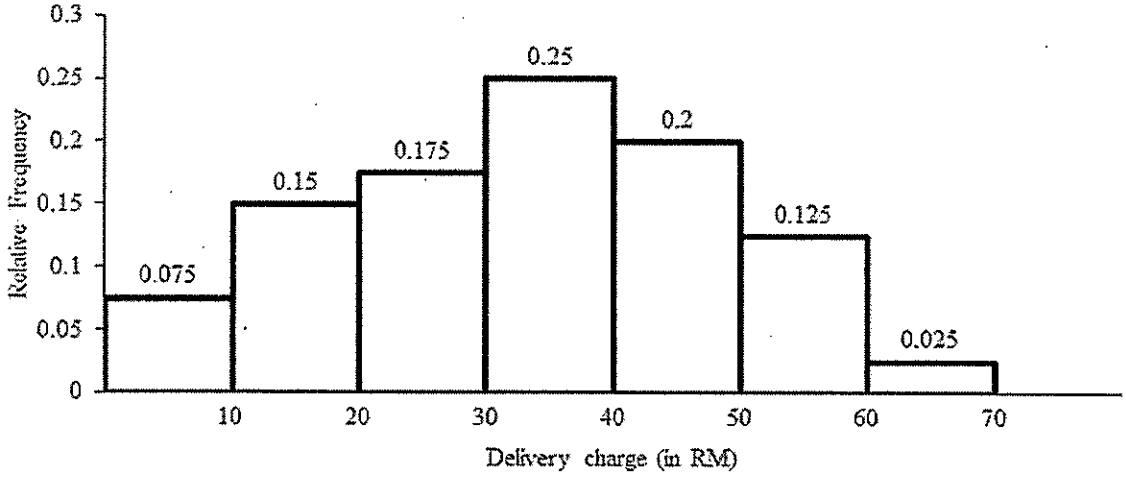
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of **3** pages excluding cover page and appendix.
2. Attempt **ALL FOUR (4)** questions. The distribution of marks for each question is given.
3. Write all your answers in the Answer Booklet provided.
4. Key formulae are given in the Appendix.

Question 1

a) The following diagram shows the relative frequency of delivery charged (in RM) on 40 consumers for the transaction made.



i. Prepare a group-data-table where the first class starts at $0 \leqslant 10$. (4 marks)

ii. Calculate the sample variance. (4 marks)

b) The following data shows the summary of time taken (in minutes) to make a quick revision for midterm test among 15 students.

25	15	10	12	13	20	30	32	q	34	21	20	11	16	15

i. Find the value of q if the average time is 20.6 minutes. (2 marks)

ii. Identify the 5-number summary. (5 marks)

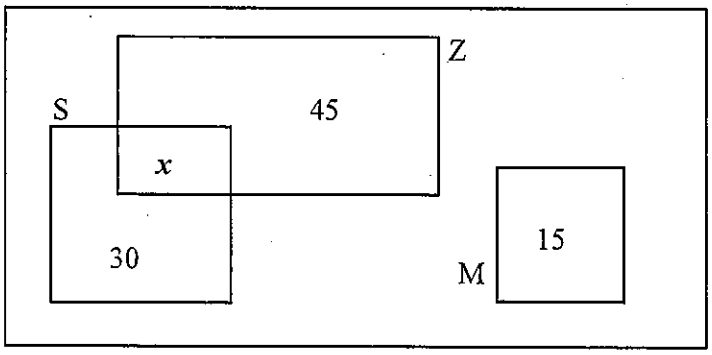
iii. From part (i) and (ii), determine the shape of the distribution. (2 marks)

iv. Find the lower and upper limit. (3 marks)

[TOTAL: 20 MARKS]

Question 2

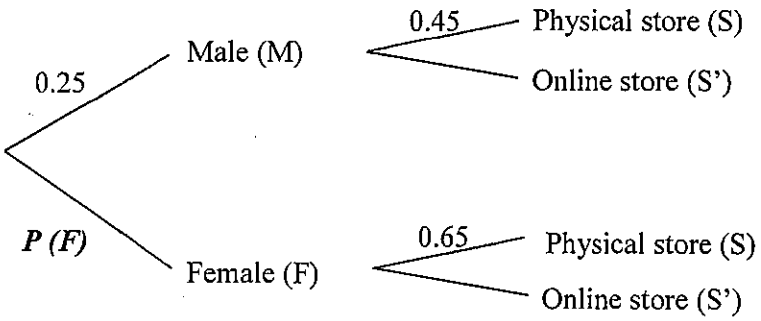
a) The Venn diagram below shows a total of 100 coffee cravers who bought the coffee for the past one month. The brands are denoted as S – Starbucks, Z – Zuss and M – McCafe. One person is selected at random.



Continued...

i. Identify the value of x. (1 mark)
ii. Find the probability of a person who buys McCafe coffee. (2 marks)
iii. Find the percentage of a person who buys either Starbucks or Zus coffee. (5 marks)
iv. Determine the probability of a person who buys a Zus coffee, given that he buys a Starbucks coffee. (5 marks)
v. What is the probability of $P(Z \cup M)$? (3 marks)
vi. Are the events Starbucks and McCafe mutually exclusive? State your reason. (2 marks)

b) A tree diagram below shows the comparison between gender and their preferences to buy groceries at physical store or to buy groceries via online store. One person is selected at random.



i. Find the value of $P(F)$. (1 mark)
ii. Find the probability of a female who prefer to buy groceries via online store. (2 marks)
iii. Determine the probability of a person who prefer to buy groceries at physical store. (4 marks)

[TOTAL: 25 MARKS]

Question 3

a) A recent survey found that, the sample mean of monthly salary of emloyees in Malaysia for year 2021 by age group of 20 to 24 years is RM1690. A random sample of 81 employees has been selected with a population standard deviation of RM1500. Test the claim at 5% significance level if the average monthly salary of employees in Malaysia for year 2021 by age group of 20 to 24 years is more than RM2000. Assume that the data is approximately normal. (10 marks)

Source: https://www.statista.com/statistics/720200/malaysia-average-monthly-salary-by-age/

Continued...

b) A survey conducted among 100 users of Cuckoo and 100 users of Coway water filter concluded that there is a difference in the average price between the two water filter. The following data summarize the results of the survey.

	Cuckoo	Coway
Population standard deviation	RM 420	RM 450
Sample mean	RM 550	RM 700

Source: https://www.cuckoonano.com/cuckoo-vs-coway/

i. Identify the null and alternative hypotheses. (2 marks)
ii. Calculate the test statistics, z. (3 marks)
iii. Find the *p-value*. (2 marks)
iv. Test the claim at 5% significance level. (3 marks)

c) A saving awareness campaign organized by EmBank shows that 75% of teenagers has opened their saving account at the early age. Based on the quarter report, only 140 out of 200 teenagers eventually opened their saving account at the early age. Does this imply that the population proportion of teenagers has been decreased? Test the claim $\alpha = 0.05$. (10 marks)

[TOTAL: 30 MARKS]

Question 4

Based on the health awareness campaign, most of the individuals consider running to get an ideal body weight. The following table shows a data from 10 individuals who perform running (in kilometer) and their body weight (in kilogram) for the past six months.

Distance run (in kilometer)	5	4	3.5	1.5	1	2	1.5	2.5	5.5	3
Body weight (in kilogram)	80	85	86	88	92	89	90	87	85	83

a) According to the table above, find the value of S_{xx}, S_{yy} and S_{xy}. (9 marks)
b) Calculate the correlation coefficient, r and interpret the answer. (4 marks)
c) Find the linear regression equation $\hat{y} = \beta_0 + \beta_1 x$. (6 marks)
d) Predict the body weight (in kilogram) if the distance run is 6.5 kilometer. (2 marks)
e) Compute the value of *SST*, *SSR* and *SSE*. (4 marks)

[TOTAL: 25 MARKS]

End of Paper.

APPENDIX – KEY FORMULA

	Ungrouped Data	Grouped Data
Mean	$\bar{x} = \frac{\sum x}{n}$ where n : sample size	$\bar{x} = \frac{\sum mf}{\sum f}$ where m : class midpoint f : class frequency
Variance	$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]$ where n : sample size	$s^2 = \frac{1}{(\sum f)-1}\left[\sum m^2 f - \frac{(\sum mf)^2}{\sum f}\right]$ where m : class midpoint f : class frequency

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $Z = \frac{x-\mu}{\sigma}$

HYPOTHESIS TESTING

Population	Mean	Proportion
1	$Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ where $\hat{p} = \frac{x}{n}$
2	$Z = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$Z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\hat{p}_p(1-\hat{p}_p)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ *where* $\hat{p}_p = \frac{x_1+x_2}{n_1+n_2}$

REGRESSION

- **Least square regression equation,** $\hat{y} = \beta_0 + \beta_1 x$

  where $\beta_1 = \frac{S_{xy}}{S_{xx}}$ and $\beta_0 = \frac{1}{n}[\sum y - (\beta_1 . \sum x)]$

- **Correlation coefficient,** $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$ where

$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$	$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$	$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$

- **Sum of Square**

$SST = SSR + SSE$	$SST = S_{yy}$	$SSR = \frac{(S_{xy})^2}{S_{xx}}$

- **Regression t-Test**

Standard Error, $s_e = \sqrt{\frac{SSE}{n-2}}$	Regression t-Test $t = \frac{\beta_1}{s_e / \sqrt{Sxx}}$	Correlation t-Test $t = r \cdot \sqrt{\frac{n-2}{1-r^2}}$